DOCUMENT RESUME

ED 167 978 CS 004 720

AUTHOR Strouss, Sara Jane
TITLE Categorizing the IEA Reading Comprehension Test Items by Ratings, Factor Analysis, Cluster Analysis, and Scalogram Analysis.
PUB DATE 78
NOTE 67p.; M.A. Thesis, University of Pittsburgh; Several pages may not reproduce due to small type

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Evaluation; Foreign Countries; Measurement Techniques; Reading Achievement; *Reading Comprehension; *Reading Research; *Reading Tests; Research Methodology; Secondary Education; *Taxonomy; *Test Items
IDENTIFIERS Barrett (T C)

ABSTRACT

One method of delineating reading comprehension is T. C. Barrett's taxonomy, which contains the categories of literal comprehension, reorganization, inferential comprehension, and evaluation. In a study devised to test whether the categories of this taxonomy are hierarchical, a reading comprehension test developed by the International Association for the Evaluation of Educational Achievement (IEA) was administered to approximately 9,000 14-year-old students in England, New Zealand, and the United States. The items of this test were classified by raters into one of the taxonomy categories and then analyzed by factor analysis, cluster analysis, and scalogram analysis. These methodologies were used to determine how each analysis explained the test items for each country and to observe the consistency across the methodologies. The findings revealed that none of the analyses strongly supported the proposed hierarchical structure of Barrett's taxonomy as measured by the IEA test items. (Tables of data are appended.) (Author/FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED167978

CATEGORIZING THE IEA READING COMPREHENSION TEST ITEMS BY RATINGS, FACTOR ANALYSIS, CLUSTER ANALYSIS, AND SCALOGRAM ANLAYSIS

By

Sara Jane Strouss

B.S., University of Pittsburgh, 1975

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sara Jane Strouss

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Submitted to the Graduate Faculty in the School of Education in partial fulfillment of the requirements for the degree of Master of Arts

University of Pittsburgh

1978

C S004 720

CATEGORIZING THE IEA READING COMPREHENSION TEST ITEMS BY RATINGS, FACTOR ANALYSIS, CLUSTER ANALYSIS, AND SCALOGRAM ANALYSIS

Sara Jane Strouss
University of Pittsburgh

ABSTRACT

One method of delineating reading comprehension is Barrett's taxonomy. This study proposed to empirically test whether the categories of Barrett's taxonomy are hierarchical. The analysis utilized a reading comprehension test developed by IEA which was given to 14-year-olds in England, New Zealand and the United States. The items of this test were classified by raters into one of the four categories of Barrett's taxonomy, as well as analyzed by factor analysis, cluster analysis, and scalogram analysis. These methodologies were used to look at how each analysis explained the test items for each country. Also, these procedures were used to observe the consistency across the methodologies. It was found that none of the analyses strongly supported the proposed hierarchical structure of Barrett's taxonomy as measured by the IEA Reading Comprehension Test items. The conflicting results of this study indicate the difficulty of validating a hierarchical structure in reading comprehension.

ACKNOWLEDGEMENTS

Through the writing of this thesis, I have come into contact with many people who have shared some of their knowledge with me. It is not possible to mention everyone, but I would like to express my appreciation to those who gave me a significant amount of guidance and encouragement.

I would like to express my deepest thanks to my thesis committee- Dr. James E. Carlson, Dr. Nancy S. Cole and Dr. Rita Bean-for their suggestions, feedback, patience, and understanding. I would like to make particular mention of Dr. Carlson's unselfish giving of his time and energy which seemed neverending.

I would like to extend a special thanks to IEA for supplying the data used in this study. The painstaking task of designing and executing the Six Subject Survey provides an important contribution to the field of education.

Also, I would like to say that the emotional support of my family and friends helped me to see through to the completion of this study. A final thanks goes to Donna Steckel for typing and retyping this thesis. Her assuming this time consuming task greatly aided me in finishing this project.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	iv
Chapter	
I. INTRODUCTION	1
II. REVIEW OF THE LITERATURE	4
The IEA Six Subject Survey	4
The Nature of Reading Comprehension	9
The Cognitive Dimensions of Barrett's Taxonomy	10
Statistical Approaches to Identifying Hierarchies	12
III. METHODS AND PROCEDURES	17
Methods	17
Procedures	17
IV. RESULTS	23
V. DISCUSSION AND CONCLUSIONS	34
REFERENCES	39
APPENDICES	
A. Layout and Computation of the Tetrachoric Correlation Coefficient	44
B. Tables of Unrotated Factor Loadings	46
C. Tables of Rotated Factor Loadings	53
D. Tree Diagrams	60

LIST OF TABLES

Table		Page
1.	Sample Sizes for Population II Schools and Students from the English Speaking Countries	7
2.	Reliability Coefficients for the Population II Reading Comprehension Test.	18
3.	Student Sample Sizes for Each Country that were Used for the Computation of the Tetrachoric Correlation Coefficient.	19
4.	A Perfect Guttman Scale	21
5.	Student Sample Sizes for Each Country Used to Form the Guttman Scales	22
6.	Classification of the IEA Reading Comprehension Test Items into the Categories of Barrett's Taxonomy as Defined by Raters	24
7.	A Comparison of the Factor Analytic Solutions for the IEA Reading Comprehension Test Items.	25
8.	The Number of Test Items for Each Factor as Defined by the Raters' Classification	26
9.	Factor Correlations for Rotated Factors from the Reading Comprehension Test Taken by the Students from England.	27
10.	Factor Correlations for Rotated Factors from the Reading Comprehension Test Taken by the Students from New Zealand.	27
11.	Factor Correlations for Rotated Factors from the Reading Comprehension Test Taken by the Students from United States.	27
12.	A Comparison of the Cluster Analytic Solutions for the IEA Reading Comprehension Test Items.	29
13.	The Number of Test Items for Each Cluster as Defined by the Raters' Classification	30

Table Page

14. Guttman Scalogram Analysis of Reading Comprehension as Defined by Barrett's Taxonomy for Students from England 32

15. Guttman Scalogram Analysis of Reading Comprehension as Defined by Barrett's Taxonomy for Students from New Zealand 32

16. Guttman Scalogram Analysis of Reading Comprehension as Defined by Barrett's Taxonomy for Students from United States 33

CHAPTER I

INTRODUCTION

Hierarchical skill categories have been proposed in a number of educational areas, and several methodological techniques have been developed to test empirically the existence of the theoretical hierarchical structures. One educational area in which a hierarchical structure has been theoretically dominant is reading, in which one hierarchical taxonomy developed to explain reading comprehension has been devised by Barrett (Barrett, 1967; Clymer, 1968; Smith and Barrett, 1974).

If Barrett's taxonomy is valid and fosters the understanding of the structure of reading comprehension skills, then it has importance for the reading teacher for use in instruction, testing, and curriculum development. The reading teacher could use the taxonomy as a framework for the instruction of the skills which a student should acquire for competency in reading comprehension. In using Barrett's taxonomy for test construction, teachers could analyze incorrect answers on tests as a means of diagnosing the students' limitations within one or more of the dimensions of reading comprehension. Then supplemental work can be provided to help students when necessary.

Barrett's taxonomy also has relevance for curriculum development. When specific requirements are defined which are necessary for reading comprehension, then organizing the materials for instruction and course sequence can follow systematically (e.g., Reading 360 program published by Ginn and Co.).

From the preceding paragraph it can be seen that a taxonomy like Barrett's could have widespread uses in education. However, little empirical testing has been done on the hierarchical structure implicit in Barrett's taxonomy. Teachers, test constructors, and curriculum developers use Barrett's taxonomy (Clymer and Ruddell, 1969; Cooke, 1971), and yet more evidence is needed to indicate empirically the presence of an underlying order to the category levels.

The characteristic of Barrett's taxonomy which was of principal interest in the present study was its hierarchical structure. The structure implies that the category levels move from easy to difficult in terms of the requirements which each category appears to address.

The purpose of this study was to consider the extent to which Barrett's hierarchical structure of reading comprehension could be empirically demonstrated in one large, multi-national test of reading comprehension by the International Association for the Evaluation of Educational Achievement (IEA), using several different statistical methodologies--ratings, factor analysis, cluster analysis, and scalogram analysis--in the analysis of the hierarchical categories in reading.

The Reading Comprehension Test, developed by IEA, was used because of its large scale application and its carefully developed sampling procedure and test construction. Its use on a multi-national level provided the opportunity also for replications across countries. It should be noted that reading comprehension per se was not of primary concern. The purpose of this study was to determine whether the Reading Comprehension Test functions in three English speaking countries tested by IEA--England, New Zealand, and the United States--in a manner

consistent with Barrett's taxonomy.[1]

Although it may be interesting to look at the item responses across population levels tested in the IEA study--10-year-olds, 14-year-olds, and end of secondary education--only 14-year-olds were considered in the present study. At this age, selection in the educational system has not yet taken place. That is, for each country in this study, 99% of the students are still in school, and therefore the controversy of comparing test results from countries with different percentages of students still in school at the end of secondary education is not relevant.

There are numerous techniques that are available to validate a category system as being a hierarchy. In this study, the hypothesized hierarchical structure of Barrett's taxonomy was investigated by the following four methods of analysis: categorization by raters, factor analysis, cluster analysis, and scalogram analysis.

---

[1] The study was limited to English speaking countries to avoid the problems of test comparability in different languages. It should be noted that students in another English speaking country were tested in reading comprehension in the original IEA study. The present study had intended to include data from this country, Scotland, in the analysis but because of difficulties encountered in obtaining the data for this country, it was not included. However, eliminating the test results of Scotland was not considered a problem, since it was thought that three was a sufficient number of countries to cross validate the proposed hierarchy.

CHAPTER II

REVIEW OF THE LITERATURE

The IEA Six Subject Survey

An empirical study in the field of education was conducted by the International Association for the Evaluation of Educational Achievement (IEA). This organization consists of twenty-two national research centers which have joined together in a cooperative effort to examine education on an international scope (Bloom, 1975). Each country has set up a national research center (hereafter referred to as National Center). Each National Center was called upon to assist in defining the sampling procedure; constructing the tests to measure achievement; and developing the various questionnaires to obtain information about the students, teachers, and schools. In other words, the National Centers' responsibility was to collaborate on the overall design for this project.

The initial effort of the IEA began in the late 1950's under the auspices of UNESCO. Many researchers from various countries were brought together with the goal of discovering whether valid achievement tests could be constructed to administer to school children from different educational systems in many countries. The first major survey measured mathematics in 12 countries. With the completion and success of this research, IEA decided in 1966 to initiate what has come to be known as the Six Subject Survey. The first phase of this endeavor took

place from 1966 to 1969 and was devoted to test construction. The second phase, which concentrated on the field testing and data analysis, was begun in 1970 and completed in 1973 (Husén, 1974). The six subjects measured in the survey were: Science, Literature, Reading Comprehension, English as a Foreign Language, French as a Foreign Language, and Civic Education.

IEA has published several reports which cover each subject separately, along with information concerning technical descriptions of the sampling process used for students, teachers, schools, test items, and questionnaire forms. In this section, those aspects of the Six Subject Survey which provide the context for this research will be described.

Target Populations

In the study three populations of students were specified. Population I ranged from ages 10 year 0 months to 10 years 11 months; Population II consisted of students 14 years 0 months to 14 years 11 months; and Population IV consisted of students in their terminal year of university preparatory programs regardless of age. There were two other populations which the National Center could have included. Population III would include students at a major terminal point in a national system somewhere between Population II and IV defined by some countries for their own use; Population IVS was a subset of Population IV that specialized in the subject being tested. The two populations just mentioned will not be discussed in the present study because of the lack of uniformity in defining these populations, and also because not all countries chose to participate in this segment of the survey.

The sampling procedure of these populations was based on those students who were actually in school at the time of the survey. There was no attempt to test individuals in an age group who were not in school. Also, the samples excluded children who were in special schools because of severe mental or physical handicaps.

In noting how the populations were defined, Populations I and II were classified by age. This was because age is uniform from country to country, while grade is not defined similarly in all countries. That is, in some countries grade is a synonym for age, while in others grade indicates a standard of achievement. Because of this second definition, the students of Populations I and II are distributed over several grades.

Population IV was not defined by age as were Populations I and II but comprised a grade group. This was determined in each country by the length of the secondary education. Therefore, Population IV may include various age groups. Also, because a selection process based on student achievement is used in some educational systems at this level, there is a wide range in the proportion of students who are still in school at this level of education (Peaker, 1975).

Sampling Plan

A master plan was recommended to achieve a stratified probability sample for each country. Within each country sampling students for this survey was accomplished in two stages. The first stage entailed the selection of schools, and the second stage involved the selection of students from these schools. The sampling procedure was designed such that every student would have a chance of being selected. Also, schools

were to be stratified by factors of relevance, i.e., sex (percentage of enrollment that is male), size, type, and region served by the school. This stratification was used to reduce sampling error (Postlethwaite, 1974).

The aim of the sampling plan was to secure at least 100 schools in each national population, and from each school to obtain approximately 30 students, both to keep the sample sizes manageable and to represent each school fairly (Peaker, 1975). There were some deviations from the recommended sampling plan, but in general, it was accepted, and the National Center for each country carried out the sampling procedure.

Because participation in the Six Subject Survey was voluntary, schools which were selected were asked if they would take part. A replacement sample of schools was also selected to be used if a school in the original selection process declined to participate in the survey. Then, after the schools were defined, the required number of students within these schools was randomly selected. Actual sample sizes for Population II schools and students from the English speaking countries can be found in Table 1. (See Postlethwaite, 1974, for a further description.)

Table 1

Sample Sizes for Population II Schools and Students from the English Speaking Countries

Country	Schools	Students
England	144	3087
New Zealand	77	1955
United States	137	3503

Test Construction

In each of the participating countries, the National Center set up a national reading committee. This committee's purpose was to review the proposed plan for the study and to submit and rev ew passages and test items.

Reading comprehension was defined as the ability to read a passage and then answer questions which refer to the passage. The passage was available for reference and re-reading.

For each population, six tryout tests were constructed. Since all population levels used similar methods of test construction, and only Population II was used in this study, Population II's test development is explained here. (See Thorndike, 1973, for further details.)

Each test for Population II contained four passages which spanned difficulty levels judged to be appropriate for students of age 12, 14, and 16 years old from a country where the average level of formal education is high. These levels were used as rough guidelines for test construction. The final test for Population II contained eight passages and 52 items. These were divided into two tests, each with a time allowance of 45 minutes.

In the original study reported by Thorndike (1973) the items of the Reading Comprehension Test (International Association for the Evaluation of Achievement, 1973) were said to have been selected with the intention of measuring the following kinds of skills:

A. Ability to determine the meaning of a word or phrase in context,
B. Ability to follow the organization of a passage and to identify antecedents and references in it,

C. Ability to answer questions that are specifically answered in the passage,
D. Ability to draw inferences from a passage about its contents, and
E. Ability to determine the writer's purpose, intent and point of view. (p. 56)

These are the kinds of skills that Barrett's taxonomy delineates. Therefore, the application of this hierarchy appears to be suitable for the present study. In the following sections, a discussion of the nature of comprehension and Barrett's taxonomy are presented.

The Nature of Reading Comprehension

Various theories have been developed concerning the nature of reading comprehension. Davis (1972) presents Chapman's three models of comprehension which illustrates the different schools of thought that have come out of the reading literature.

1. The uncorrelated, or isolated, skill theory, which postulates that comprehension is made up of a set of skills (or mental processes) that are learned and used independently, and in any order.
2. The global-skill theory, which postulates that comprehension is a unitary ability that, in combination with errors of measurement alone, accounts for all of the variance of measurements of comprehension in reading.
3. The hierarchical skills theory, which postulates that comprehension is made up of separate but correlated skills and that these differ in complexity because the more complex include all or parts of the simpler, or more basic, ones. (pp. 670-671)

The model that currently predominates in the literature in the area of reading comprehension is that reading comprehension is a complex, unified, continuous activity which does not fall naturally into discrete measurable skills (Traxler, 1970). However, because there is little empirical evidence, this global-skill model does not negate the possibility that reading comprehension is a domain composed of hierarchical components. Indeed, the current trend towards criterion-

referenced testing and skill classification systems indicate some support for the hierarchical skills theory (Spache and Spache, 1973).

The Cognitive Dimensions of Barrett's Taxonomy

Barrett's taxonomy was developed in accordance with the hierarchical skills model. This taxonomy was a reaction to what Barrett stated to be two misconceptions in the measurement and instruction of reading comprehension. First, that reading comprehension is a single unitary skill, and second, that comprehension in reading contains many separate skills which makes it unmanageable (Clymer, 1968; Smith and Barrett, 1974).

The five major categories in this taxonomy are: literal comprehension, reorganization, inferential comprehension, evaluation, and appreciation. Although this taxonomy includes a category for appreciation, this study did not encompass this aspect of the taxonomy. Reading comprehension, as measured by the reading test in this study, was not developed to evaluate appreciation. The Six Subject Survey measured literature as a separate subject, and appreciation was evaluated by that test.

Also, Barrett's taxonomy does not deal specifically with determining the meaning of a word or phrase in context. However, questions of a vocabulary nature can be classified by the taxonomy. For example, the following question taken from Section C of the IEA Reading Comprehension Test (International Association for the Evaluation of Educational Achievement, 1973) has been classified as inferential comprehension:

"Authority" as used in line 6 of the above article, means

A. traditional wisdom.
B. scientific analysis.
C. inductively determined fact.
D. superstition.

Another point that should be made is that the reorganization category has been taken out of the current breakdown of the taxonomy (Smith and Barrett, 1974). However, there has not been any documented reasoning for deleting this category. Logically, this category appears to have relevance in defining reading comprehension, and therefore, has been included in the present study to observe whether it is pertinent to include it in the proposed hierarchy.

Clymer (1968) gives a detailed description of the categories and subcategories of Barrett's taxonomy. A summary of the taxonomy is included here.

Literal comprehension. The primary focus of this category is centered on the ideas and information which are explicitly stated in the selection. An example of this would be to recall certain details of what was read, such as the location of where the story took place.

Reorganization. In this category the ideas or information which are explicitly stated in the selection need to be analyzed, synthesized, and/or organized by the student. One way to measure this could be to require the student to categorize events which took place in the passage.

Inferential comprehension. The information which is explicitly stated in the passage is used by the student to make conjectures and hypotheses (e.g., inferring relationships which could cause a particular outcome).

Evaluation. A student demonstrates understanding in this category by indicating that an evaluative judgement was made, such as judging whether the author's conclusions follow from the premises put forth in the passage.

The categories of this classification system were developed such that they are ordered from easy to difficult in terms of what appears to be required by the student to demonstrate the attainment of the accompanying skills. An advantage of this method of defining reading comprehension is the systematic way of encompassing each educational objective. This taxonomy could provide a means of communication for the reading teacher, as well as permitting the student to realize the sequence of skills which lead to reading comprehension.

Statistical Approaches to Identifying Hierarchies

Hierarchical structures in learning theory have provided a systematic method for classification of educational goals. Validation of these hierarchical structures must be done through application of statistical methodologies in order to give empirical evidence of logically sequenced steps to an educational objective.

Many techniques of various levels of difficulty have been applied to previously proposed hierarchical classification schemes. However, no one statistical method has yielded conclusive evidence for testing hierarchies. Therefore, in this study several techniques were used, and in this section an overview of these statistical approaches to identifying a hierarchy will be discussed. The implications of applying several methodologies on the same data enables the researcher to observe the consistency of the resultant structures found in the analyses.

That is, does the hypothesized hierarchical structure of Barrett's taxonomy exist in the samples used in this study for factor analysis, cluster analysis, and scalogram analysis?

Factor analysis, cluster analysis, and scalogram analysis are three methodologies which attempt to explain test items in terms of an underlying structure. In general, the object of factor analysis is that of data reduction. That is, given a matrix of correlation coefficients, the technique of factor analysis enables the researcher to observe the possible existence of some underlying relationships in the data. This is done so as to reduce the number of factors and still account for the observed interrelationships in the data.

An example of factor analysis as applied to Barrett's taxonomy can be found in a study which investigated two categories of Barrett's taxonomy--literal and inferential comprehension--conducted by Pettit and Cockriel (1974). Through principal component analysis, literal and inferential comprehension were found to be two separate distinguishable skills. Maximum likelihood factor analysis yielded a very similar factor structure to that found through the principal component analysis. This lends some support to the hypothesis which states that there are at least two categories that define reading comprehension. Although only two of Barrett's categories were used in the Pettit and Cockriel study, the analysis indicated that a hierarchical structure does exist, since items which primarily loaded on the literal comprehension factor, and were considered a separate dimension in reading, did have moderate loadings on the inferential comprehension factor. This means that

literal comprehension items were related to inferential comprehension items, a relation which always occurs if there is a hierarchy, but which in itself is not a sufficient condition for the existence of a hierarchy. It also suggests that the understanding of literal comprehension precedes that of inferential comprehension. Also, the intercorrelations between the literal and inferential comprehension categories indicated that these categories tended to be correlated. When correlated factors exist, then, by definition, an oblique rotation in factor analysis is done. This lends support to the present study's use of an oblique rotation.

Although general factor analysis procedures address the structure in a set of data, they were not designed directly to identify hierarchical structures. One must further interpret the factors and factor loadings found. In observing the intercorrelations of the factors defined by factor analysis, a direct investigation of a hierarchical structure can be looked at by simplex theory. In this theory a perfect simplex is characterized by the larger correlations being found next to the main diagonal, and taper off as one goes to the upper right and lower left of the matrix of intercorrelations (Guttman, 1954). If this occurs then the factors are said to define an ordinal scale that goes from least complex to most complex. If the factors do not form a perfect simplex, but the intercorrelations approach a simplical structure, then the factors are said to form a quasi-simplex. If a simplex or a quasi-simplex exists, then there is evidence that the factors support the hypothesized hierarchy of complexity.

Cluster analysis is designed to create classifications by forming groups of relatively homogeneous entities. This method of analysis looks at the entities, in terms of their similarity, to form the clusters. One procedure which has relevance for the present study is a hierarchical clustering scheme. Clusters are defined by similarities or distances between the entities. One particular method of defining distance is by the complete linkage algorithm which defines the distance from an entity in the first cluster to an entity in the second cluster as the longest distance from an entity in the first cluster to an entity in the second cluster. The order in which the items are clustered is determined by the strength of association. Strength of association is defined as either similarities (e.g., correlation coefficients) or distances (e.g., Euclidean distances). When similarities are used, the association becomes stronger as the value of the similarity between items becomes larger. The opposite is true when the strength of association is a distance matrix--the association becomes stronger as the distance between items becomes smaller.

Scalogram analysis is a technique which rank orders items along a continuum in terms of the complexity of the items. This scaling procedure is a direct method of investigating whether a hypothesized hierarchy is tenable, and if it is, then knowing a person's scale score, one also knows the response pattern (i.e., the items that the person got right and items the person got wrong). A measure of the extent to which a person's response pattern can be reproduced based on a scale score is the coefficient of reproducibility. Although this is the principal test for scalability, it is not a sufficient criterion to confirm that the items are scalable (Guttman, 1950). Another measure

to consider is the minimum marginal reproducibility. This is the minimum coefficient of reproducibility that could have occurred for the scale given the proportion of respondents passing and failing each of the items. Looking at the difference between the coefficient of reproducibility and the minimum marginal reproducibility provides an indication of the extent that the coefficient of reproducibility is effected by the response patterns as opposed to the inherent cumulative interrelation of the items used. This difference is known as the percent improvement. The last measure used in scalogram analysis is the coefficient of scalability. This coefficient is used to indicate if the scale is truly unidimensional and cumulative.

If a hypothesized hierarchy is found to be valid in each of the methodologies presented in this section--factor analysis, simplex theory, cluster analysis, and scalogram analysis--then it seems reasonable to conclude that it is indeed a hierarchy. Similarly, the lack of a hierarchical structure in all four methodologies would strongly suggest the lack of such a structure in the reading items of the IEA test and raise the possibility of the lack of hierarchical structure in many tests similar in content to the IEA. On the other hand, conflicting results among the four procedures would be instructive for examining the procedures themselves and the structures to which they are sensitive as well as for examining the relationships among the reading categories.

CHAPTER III

METHODS AND PROCEDURES

Methods

Subjects. The source of data for this study was the Population II Reading Comprehension item responses from those students who were sampled from England, New Zealand, and the United States.

Instruments. The Reading Comprehension Test, as described earlier, focused on the cognitive aspects of reading a passage which is followed by multiple choice questions referring to the content of the passage. The two forms of this test--E2RC and E2RD--were analyzed as one test, by categorizing each item with respect to Barrett's taxonomy.

Two measures of reliability for the Population II Reading Comprehension Test from each country in this study are reported in Table 2 (Thorndike, 1973). The Kuder-Richardson Formula 20 (KR-20) represents the internal consistency within the total test for each country. The alternate form reliability coefficient is a correlation between the two separately timed subtests which has been corrected by the Spearman-Brown Prophecy Formula to give an estimate of the reliability of the total test.

Procedures

Items on the Reading Comprehension Test were categorized into one of four levels of Barrett's taxonomy--literal comprehension, reorganization, inferential comprehension, and evaluation. This was done

Table 2

Reliability Coefficients for the Population II Reading Comprehension Test

Country	KR-20	Alternate Form
England	.887	.864
New Zealand	.875	.857
United States	.885	.851

by an associate professor in Educational Research Methodology, University of Pittsburgh, who has had experience in the classification of test items, and an assistant professor in the Department of Language Communications, University of Pittsburgh, who has had experience with the application of Barrett's taxonomy. Items were classified separately then discrepant classifications were discussed. With discussion, the two raters agreed on the classification of all items.

Based on the categorization by the raters, the items which formed each level of Barrett's taxonomy were defined. This provided the theoretical classification that the remaining analysis utilized for comparison of the skills which the items appeared to be measuring.

The hierarchical nature of this taxonomy is tenable if there is an order to the skills needed for each successive category. That is, the categories were ranked from least complex, in terms of skills required, to the most complex group of necessary skills. Each successive category in this theoretical structure was assumed to be built upon the knowledge of the prior categories, which in turn leads to the educational objective of reading comprehension. The following order should have resulted if Barrett's taxonomy, as previously defined, is hierarchical:

literal comprehension, reorganization, inferential comprehension, and evaluation.

For each person who took the Reading Comprehension Test at the Population II level (see Table 3 for sample sizes), the items were scored as correct or incorrect. A value of one was assigned to correct answers and a value of zero for incorrect answers. Table 3 reports the sample sizes used in this study, which differ from the sample sizes reported in Table 1 because of inconsistencies in the IEA literature for the reported sample sizes.

Table 3

Student Sample Sizes for Each Country that were Used for the Computation of the Tetrachoric Correlation Coefficients

Country	Students
England	3284
New Zealand	1995
United States	3469

After reviewing the measures of association available to explain relationships of categorical data scored dichotomously (Sokal and Sneath, 1963), the tetrachoric correlation coefficient was found to be appropriate for the following analyses. This conclusion was based on the assumption that the variable being measured is normally distributed. A tetrachoric correlation coefficient was computed between items which yielded a 52x52 matrix for each country. The procedure that the SPSS computer program package (Nie, Hull, Jenkins, Steinbrenner and Bent, 1975) used to compute these correlations can be found in Appendix A.

One way this study investigated a hierarchical structure was by factor analyzing the test items, using the maximum likelihood method calculated by the BMDP4M program (Dixon, 1975). If a hierarchy exists, then the factors are correlated, and a direct quartimim oblique rotation was done to determine if this relationship occurred.

A hierarchical clustering procedure was also used. This is a systematic method of bringing in items which are similar, to form a cluster. Similarity is defined by forming a similarity-dissimilarity matrix (in this case the matrix of tetrachoric correlation coefficients) which is the strength of association between items (Blashfield, 1976).

The particular clustering technique that was applied was an agglomerative method computed by BMDP1M (Dixon, 1975). This begins with each item as a separate cluster and proceeds until all items belong to one cluster. Two techniques were considered to link the items together to form the clusters--single linkage and complete linkage. Single linkage defines a cluster such that every entity is more similar to at least one entity in the cluster than to any other entity of another cluster. Complete linkage defines a cluster such that each entity is more similar to all entities in the given cluster than to all entities in any other cluster. Because single linkage has the tendency to chain (the clusters tend to resemble long chains when plotted in a multidimensional space) (Blashfield, 1976), this method was not included in the analysis. Complete linkage clustering was completed for each sample in this study.

Based on the raters' classification of the items into the levels of Barrett's taxonomy, the SPSS program of scalogram analysis was used to see if the categories formed a Guttman scale (Nie et al., 1975).

This means that the taxonomy levels were scalable and hierarchical if the following order resulted: literal comprehension (1), reorganization (2), inferential comprehension (3), and evaluation (4). An example of a perfect Guttman scale can be found in Table 4.

Table 4

A Perfect Guttman Scale

Subject	Category 1	2	3	4	Score
1	1	1	1	1	4
2	1	1	1	0	3
3	1	1	0	0	2
4	1	0	0	0	1
5	0	0	0	0	0

In order to apply scalogram analysis, an additional coding process was used. If 80% of the items within a category were correct, a value of one was assigned to that category. Otherwise, the category was assigned a value of zero. The 80% criterion was based on what Graham (1966) considered as passing a "contrived" item--a contrived item being formed from several items which are necessary to test adequately one category. This process was done for each category (i.e., each "contrived" item) for all students. This resulted in each student having four values of either zeros or ones. Those students who did not respond to all test items were deleted from this coding process because of complications in forming the "contrived" items when missing data were involved. Sample sizes for this procedure can be found in Table 5.

Table 5

Student Sample Sizes for Each Country Used to Form the Guttman Scales

Country	Students
England	2256
New Zealand	1604
United States	2998

If Barrett's taxonomy is scalable in the hypothesized hierarchy, then a score of 4 indicates that all categories were passed. A score of 2, then, implies that the student passed literal comprehension and reorganization but failed inferential comprehension and evaluation. In other words, from the score a student received, one could reproduce which categories were passed and which were not passed.

Looking at the results of each analysis in this study, a comparison within each method was made of each country's results. This comparison was used to view the consistency of the analysis when applied to different samples--a means of cross validation. The results were also summarized across techniques, and this comparison was used to determine if the structure of the category levels was hierarchical for all the methods used in this study.

CHAPTER IV

RESULTS

The Population II IEA Reading Comprehension Test items were analyzed for students from England, New Zealand, and the United States. In reporting the results in this study, form E2RC will be referred to as items 1 through 26 and form E2RD will be referred to as items 27 through 52.

The classification of the test items into the categories of Barrett's taxonomy, as defined by the raters, can be found in Table 6. It can be seen from the table that the raters found the Reading Comprehension Test to be predominantly measuring what Barrett's taxonomy has defined as literal and inferential comprehension. It should be noted that only one item was considered to be measuring Barrett's category of evaluation, and only three items were categorized as measuring reorganization.

The results of the factor analyses are reported in Tables 7 through 11. It was found that the total variance that was explained by the four factor solution for the English sample was 30%; in the New Zealand sample the percent variance explained was 30.7%; and for the United States sample, the variance explained was 30.5%. The unrotated factor loadings are reported in Appendix B. The tables in this Appendix show the percent of variance explained by each factor.

Table 6

Classification of the IEA Reading Comprehension Test Items into the Categories of Barrett's Taxonomy as Defined by Raters

Category	Item Numbers
Literal Comprehension	1, 8, 9, 15, 18, 21, 22, 25, 30, 32, 34, 42, 47, 48, 50
Reorganization	19, 20, 46
Inferential Comprehension	2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 16, 17, 23, 24, 26, 27, 28, 29, 31, 33, 35, 36, 37, 38, 39, 40, 41, 43, 44, 49, 51, 52
Evaluation	45

The factor pattern matrices were rotated to obtain oblique solutions. Table 7 provides a means for comparing across samples the items which load on each factor. (Complete tables of the rotated pattern matrices can be found in Appendix C.) A cut off value of .300 was used to define the items which load on each factor. It can be seen that some items on the Reading Comprehension Test did not contribute to any of the factors. Specifically, in England 12 items did not load on any of the specified factors, for New Zealand 14 of the items were not represented in the four factor solution, and in the United States 11 reading comprehension items did not load on any of the specified factors. After reviewing the factor loadings, it was found that two items loaded on two factors in the English sample, for the New Zealand sample five items loaded on two factors, and three items loaded on two factors in the sample from the United States.

Table 7

A Comparison of the Factor Analytic Solutions for the IEA Reading Comprehension Test Items

Factor 1			Factor 2			Factor 3			Factor 4		
EN	NZ	US	EN	NZ	US	EN	NZ	US	EN	NZ	US
9			2		2*	1**				1**	1
10**	10**	10**			3**	2	2			2	
12**	12**	12**	4*	4*	4**	3*				3	
					5	5**				8	
			6	6	6*		13			9**	
			7**	7	7**		15			14	
			8		8*		16			21*	
					9		17			23	
			11				18			30	
			13**		13**		19	19		33*	
			14		14**		20	20		39*	
			15*	15	15**	21				43	
			16**	16	16**		28*			51	
					22		29				
			23*	23*	23**		32				
			24**	24**	24**	39*	39				
			25**	25*	25*		40				
					27**		42*	42*			
			28*		28*	43		43			
					29			45			
			30		30		46*				
			31*		31*		50*	50			
			32*		32		52	52			
			33		33**						
			34	34	34*						
			35*		35*						
			36*		36*						
				37							
			38**								
					39						
			40*		40						
			42**		42						
			43		43						
			44*								
			45*		45						
			46		46						
					47						
			48*	48	48*						
			49								
			50								
			52*								

Factor loading between .300 and .399
* Factor loading between .400 and .499
** Factor loading .500 or greater

In trying to define the factors in terms of Barrett's taxonomy, how the items factored with respect to the raters' categories can be found in Table 8. As noted previously, a few items in each sample loaded on two factors, and therefore have been counted for each factor. It can be seen that the derived factors do not consistently conform to Barrett's taxonomy.

Table 8

The Number of Test Items for Each Factor as Defined by the Raters' Classification

Barrett's Categories	Factor 1			Factor 2			Factor 3			Factor 4		
	EN	NZ	US	EN	NZ	US	EN	NZ	US	EN	NZ	US
Literal Comprehension(15)	1			9	4	11	2	5	2		5	1
Reorganization(3)				1		1		3	2			
Inferential Comprehension(33)	2	2	2	21	7	21	5	9	2		8	
Evaluation(1)				1		1			1			

The intercorrelations of the factors for England, New Zealand, and the United States appear in Tables 9, 10, and 11, respectively. From these tables it can be seen that there is a tendency for the correlations to become smaller moving away from the main diagonal. This indicates some support for a quasi-simplex which says that the hypothesized order of factors from least complex to most complex has some validity. However, because of the ambiguity in defining the factors within each sample, as well as the lack of consistency across samples, the factor correlations cannot be interpreted as supporting Barrett's taxonomy as a hierarchy.

Table 9

Factor Correlations for Rotated Factors from the Reading Comprehension Test Taken by the Students from England

	Factor			
	1	2	3	4
Factor 1	1.000			
Factor 2	.612	1.000		
Factor 3	.331	.456	1.000	
Factor 4	-.087	.082	.019	1.000

Table 10

Factor Correlations for Rotated Factors from the Reading Comprehension Test Taken by the Students from New Zealand

	Factor			
	1	2	3	4
Factor 1	1.000			
Factor 2	.390	1.000		
Factor 3	.383	.541	1.000	
Factor 4	.356	.353	.325	1.000

Table 11

Factor Correlations for Rotated Factors from the Reading Comprehension Test Taken by the Students from United States

	Factor			
	1	2	3	4
Factor 1	1.000			
Factor 2	.579	1.000		
Factor 3	.253	.450	1.000	
Factor 4	.174	.212	.121	1.000

Using the complete linkage cluster analytic solutions provided another method of looking at the proposed hierarchical pattern of the IEA Reading Comprehension Test items. The ordering of each item was based on the item's similarity with the other items. That is, those items most similar were joined together first, and this continued until the least similar items were joined.

In reporting the breakdown of the cluster analyses, the number of clusters is subjective. Therefore, in order to be consistent with the other analyses in this study, four clusters were chosen for each sample based on the tree diagrams in Appendix D. These tree diagrams are a graphic method of displaying the hierarchical clusters that resulted when the clustering procedure operated on the similarity matrix (Duran and Odell, 1974). The resulting breakdown of the IEA items for each sample is given in Table 12. This table also presents the strength of association for each cluster which is the similarity of the items within each cluster expressed as a correlation coefficient that has been scaled from 0 to 100.

As a means of linking the cluster analytic solutions with the raters' classification, Table 13 delineates the number of items in each cluster that fell into each level of Barrett's taxonomy according to the raters' categorization.

Table 12

A Comparison of the Cluster Analytic Solutions for the IEA Reading Comprehension Test Items

Cluster 1			Cluster 2			Cluster 3			Cluster 4		
EN	NZ	US	EN	NZ	US	EN	NZ	US	EN	NZ	US
1	1	1		4		4			17	17	17
2	2	2		5		5			18		
3	3	3		6	6	6			19	19	19
		4		7		7			20	20	20
		5			10	10	10		26	26	
		7	11		11	12	12				37
8	8	8			12			18			38
9	9	9		13				21	51	51	
	11		15	15			22	22			
13		13	16	16				26			
14	14	14	25	25	25		47				
		15	29				49				
		16	33				50	50			
	18		34	34				51			
21	21		37	37							
22				38							
23	23	23	41		41						
24	24	24			43						
27	27	27	45	45	45						
28	28	28	47								
	29	29		48							
30	30	30	49		49						
31	31	31	50								
32	32	32	52	52	52						
	33	33									
		34									
35	35	35									
36	36	36									
38											
39	39	39									
40	40	40									
	41										
42	42	42									
43	43										
44	44	44									
46	46	46									
		47									
48		48									
Strength of Association for Each Cluster											
58.5	52.9	57.1	53.2	55.4	53.9	52.6	55.4	53.8	49.1	45.2	49.3

Table 13

The Number of Test Items for Each Cluster as Defined by the Raters' Classification

Barrett's Categories	Cluster 1			Cluster 2			Cluster 3			Cluster 4		
	EN	NZ	US	EN	NZ	US	EN	NZ	US	EN	NZ	US
Literal Comprehension(15)	9	8	10	5	4	1		3	4	1		
Reorganization(3)	1	1	1							2	2	2
Inferential Comprehension(33)	16	18	20	8	9	8	6	3	2	3	3	3
Evaluation(1)				1	1	1						

The final method for investigating the hypothesized hierarchical structure of Barrett's taxonomy was Guttman's scalogram analysis. The SPSS program, which computed the Guttman scales, reported the frequency that respondents passed each category of the scale. In order to compare the results with a perfect Guttman scale (see Table 4), these frequencies were converted to the proportion of respondents who passed each category based on the total number of respondents with each score.

The SPSS program also allowed the categories to be scaled automatically according to the percentage of respondents that failed each contrived item. That is, the contrived item which was considered most difficult was the one with the largest proportion of respondents to have failed it. This procedure continued until the least difficult contrived item was defined, that being the contrived item with the smallest proportion of respondents to have failed it. In keeping the results consistent with the model in Table 4, the categories are reported from least difficult to most difficult.

Table 14 reports the results of the Guttman scalogram analysis for the sample from England; Table 15 presents the results of this analysis for the sample from New Zealand; and Table 16 gives the results of this analysis for the sample from the United States. The resultant hierarchy shown in each of these tables did not agree with the hypothesized hierarchy. The differences lies with the categories of literal and inferential comprehension being interchanged.

In looking at the resultant hierarchy, the statistical analyses included the coefficient of reproducibility, the minimum marginal reproducibility, the percent improvement, and the coefficient of scalability. A coefficient of reproducibility of 0.9 or greater (based on the range of 0 to 1) has been used as a general guideline for considering a scale as a valid hierarchy. It can be seen from Tables 14, 15 and 16 that in the strict sense, this criterion was not met. However, the coefficient of reproducibility for each sample was close, and therefore, not meeting this criterion does not necessarily negate the resultant hierarchy.

The coefficient of scalability is based on the ratio of the percent improvement to the difference between 1 and the minimum marginal reproducibility. The range of the coefficient of scalability is 0 to 1. The value of this coefficient should be above .6, if the scale is truly unidimensional and cumulative. Based on this criterion, the resultant scales from New Zealand and the United States supported this requirement. However, England's resultant scale yielded a coefficient of scalability slightly less than .6.

Table 14

Guttman Scalogram Analysis of Reading Comprehension as Defined by Barrett's Taxonomy for Students from England

	Category			
Score	Inferential Comprehension	Reorganization	Literal Comprehension	Evaluation
4	1	1	1	1
3	1	.864	.755	.380
2	1	.629	.192	.179
1	1	0	.001	0
0	0	0	0	0

Coefficient of Reproducibility = .888
Minimum Marginal Reproducibility = .730
Percent Improvement = .158
Coefficient of Scalability = .585

Table 15

Guttman Scalogram Analysis of Reading Comprehension as Defined by Barrett's Taxonomy for Students from New Zealand

	Category			
Score	Inferential Comprehension	Reorganization	Literal Comprehension	Evaluation
4	1	1	1	1
3	1	.896	.852	.252
2	1	.576	.309	.115
1	1	0	0	0
0	0	0	0	0

Coefficient of Reproducibility = .893
Minimum Marginal Reproducibility = .713
Percent Improvement = .180
Coefficient of Scalability = .628

Table 16

Guttman Scalogram Analysis of Reading Comprehension as Defined by Barrett's Taxonomy for Students from United States

Score	Category			
	Inferential Comprehension	Reorganization	Literal Comprehension	Evaluation
4	1	1	1	1
3	1	.806	.705	.489
2	1	.632	.180	.190
1	1	.002	0	0
0	0	0	0	0

Coefficient of Reproducibility = .880
Minimum Marginal Reproducibility = .684
Percent Improvement = .196
Coefficient of Scalability = .622

CHAPTER V

DISCUSSION AND CONCLUSIONS

Several statistical models for validating a taxonomy of reading comprehension have yielded conflicting yet instructive results. This study has attempted to use these models in two ways. One way was to look at how each analysis explained the test items for each sample. The second way was to observe the consistency across the methodologies.

In accordance with this second procedure, the following is an example of the ambiguity encountered in this study. As mentioned previously, the rater's classification primarily emphasized that the IEA Reading Comprehension Test was measuring Barrett's categories of literal and inferential comprehension. Since the primary means of comparison were these raters' classifications and essentially only two categories resulted from this analysis, purposing four categories may seem unfounded. However, when the factor analyses and cluster analyses were done independently of these racings before being looked at in terms of these ratings, there appeared to be more than two categories needed to explain what was underlying the test items. One interpretation of the inconsistency that occurred could lie in the difficulty of operationalizing Barrett's taxonomy when translating what the test items were measuring.

Another problem was encountered when the amount of variance which the factor analytic solutions explained was found to be approximately 30% for each sample. This means that 70% of the variance remained unexplained suggesting that more than four dimensions underlie

these reading comprehension items. This is in contradiction to Barrett's hypothesis of four categories to classify reading comprehension. This also puts Barrett's deletion of the reorganization category in a more questionable light.

If what the reading comprehension test is measuring is unidimensional, one large factor would have resulted in the factor analytic solutions. However, the largest factor in the three samples ranged from 16% to 21% of explained variance which does not support the proposal that this test is measuring a unidimensional domain. On the other hand, if the criterion that the eigenvalues have a value of one or greater was used, it was found that 12 factors would be needed to explain the underlying dimensions of the test items.

If there is consistency across the statistical analyses, then there should be more than four clusters to explain the underlying dimensions of the test items. Because of the subjective nature of the number of cluster groups that exist, it is possible that the clusters should be broken down into smaller groups consequently yielding more clusters.

The results of the Guttman scalogram analysis questioned the order of Barrett's categories. Using only the Guttman scaling procedure, the reordering of the literal and inferential comprehension categories might have been strongly suggested. However, the other analyses were less clear on this point, leaving a more questionable view of the order of Barrett's taxonomy.

Since the proposed hierarchical system of defining reading comprehension was not supported, other explanations were sought to try

to explain the relationship of the test items found through the various analyses. For example, the items were examined according to the paragraph on which they were based. However, this type of interpretation did not provide any further information that could be used to resolve why the items followed the patterns that resulted. This type of analysis brings up a topic which has been a concern in the area of reading. In writing items to determine the understanding of a passage, the item should be dependent on reading the passage in order to answer it correctly (Pyrczak and Axelrod, 1976; Tuinman, 1973-74). If the item is not dependent on the passage then using this as a means of evaluating a student's level of comprehension would be misleading.

The resultant hierarchy of the scalogram analysis provides an interesting view of reading comprehension which suggests that the inferential comprehension items on the IEA test were easier than the literal comprehension items. Relating this to the previous paragraph on passage dependence, one interpretation of this resultant hierarchy could be that if you only scan the passage, it is harder to get the specific items correct and easier to get the more general type of items correct.

An issue which has presented itself when discussing a taxonomy of reading comprehension is whether reading comprehension can be mastered or is it content specific. Mastering reading comprehension would mean that the skills which define reading comprehension could be applied to any subject matter. But if reading comprehension is in part content specific, then another question arises about the extent to which the common elements provide a framework for reading comprehension.

Suggestions for Further Research

In order to answer the questions brought up in the preceding paragraph, it would be required to design a study which analyzes types of subject matter. The skills needed to comprehend the content of each type of subject matter would be defined. The comparison of these skills across subject matter would indicate the amount of communality across the content of subject matter. The more elements that the subject matter has in common, then the less likely that reading comprehension would be considered to be content specific.

A procedure to more directly examine the measurement of reading comprehension as defined by Barrett's taxonomy would be to construct a test to specifically measure the skills within each category of this taxonomy. Using this instrument as a measure of reading comprehension, four scores would be derived from each student's answers which would be measuring the four factors of Barrett's taxonomy. The intercorrelation of these factors could be used for investigating if a simplex structure exists, and the results could then be interpreted in terms of Barrett's taxonomy. The other procedures in this study could also be used with this instrument that specifically utilized Barrett's categories in its construction.

Although this study tried to encompass procedures which are seldom, if ever, applied to the same data, other techniques of analysis could also be investigated.

A technique which looks at the problem of studying the similarities and differences in factor structures among different groups has been examined by Jöreskog (1971). It would have been advantageous to

use this technique in analyzing the data in the present study. However, because of the size of the data (i.e., the large number of variables), the computer program, SIFASP (van Thillo and Jöreskog, 1970), was not applicable. A program, which this researcher did not have access to but which has the capacity to handle larger amounts of data, is known as COFAMM (Sörbom and Jöreskog, 1976). Future studies of this nature should utilize the type of rigorous analysis which SIFASP and COFAMM employ.

Another scaling technique which is an extension of the Guttman scalogram analysis has been developed by Lingoes (1963). Lingoes' multiple scalogram analysis (MSA) is a method of testing a Guttman model rather than imposing this type of model. Also, when multiple scales are required to explain the data, MSA has the capacity of fitting the scales as opposed to rejecting the items as not being scalable. Unfortunately, in the present study the computer program for MSA could not be modified to handle the large sample sizes.

REFERENCES

Barrett, T. C. Goals of the reading program: The basis for evaluation. In T. C. Barrett (Ed.), The evaluation of children's reading achievement. Newark, DE: International Reading Association, 1967.

Blashfield, R. K. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. Psychological Bulletin, 1976, 83, 377-388.

Bloom, B. S. Implications of the IEA studies for curriculum and instruction. In A. C. Purves and D. U. Levine (Eds.), Educational policy and international assessment. Berkeley, CA: McCutchan Publishing Corporation, 1975.

Clymer, T. What is reading? Some current concepts. 67th Yearbook of National Society for the Study of Education, 1968, 7-29.

Clymer, T. and Ruddell, R. B. How it is nowadays. Boston: Ginn Publishing Company, 1969.

Cooke, D. A. An analysis of reading comprehension questions in basal reading series according to the Barrett taxonomy. (Doctoral dissertation, Cornell University, 1970). Dissertation Abstracts International, 1971, 31, 6467A. (University Microfilms No. 71-72, 124.)

Davis, F. B. Psychometric research on comprehension in reading. Reading Research Quarterly, 1972, 7, 628-678.

Dixon, W. J. (Ed.). BMDP biomedical computer programs. Los Angeles: University of California Press, 1975.

Duran, B. S. and Odell, P. L. Cluster analysis: A survey. In Beckmann and H. P. Künzi (Eds.), Lecture notes on economics and mathematical systems (Vol. 100). Heidelberg: Springer-Verlag, 1974.

Graham, G. T. Sequentially scaled achievement tests: Construction methodology and evaluation procedures. (Unpublished doctoral dissertation, University of Pittsburgh, 1966.)

Guttman, L. A basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star and J. A. Clausen (Eds.), Measurement and prediction (Vol. 4). Princeton: Princeton University Press, 1950.

Guttman, L. A new approach to factor analysis: The radax. In P. F. Lazarsfeld (Ed.), Mathematical thinking in the social sciences. Glencoe, Ill.: The Free Press, 1954.

Husén, T. Introduction to the reviews of three studies of the International Association for the Evaluation of Educational Achievement (IEA). American Educational Research Journal, 1974, 11, 407-408.

International Association for the Evaluation of Educational Achievement. IEA Six-Subject Survey Instruments: Reading comprehension tests. Sweden, 1973. (ERIC Document Reproduction Service No. ED 102 172)

Jöreskog, K. G. Simultaneous factor analysis in several populations. Psychometrika, 1971, 36, 409-426.

Lingoes, J. C. Multiple scalogram analysis: A set-theoretic model for analyzing dichotomous items. Educational and Psychological Measurement, 1963, 23, 501-524.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. and Bent, D. H. SPSS: Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill, 1975.

Passow, A. H., Noah, H., Echstein, M. A. and Malles, J. International studies in evaluation VII: The national case study: An empirical comparative study of twenty-one educational systems. Stockholm: Almqvist and Wiksell, 1975.

Peaker, G. F. International studies in evaluation VIII: An empirical study of education in twenty-one countries: A technical report. Stockholm: Almqvist and Wiksell, 1975.

Pettit, N. T. and Cockriel, I. W. A factor study of the literal reading comprehension test and the inferential reading comprehension test. Journal of Reading Behavior, 1974, 6, 63-75.

Postlethwaite, T. N. Target populations, sampling, instrument construction and analysis procedures. Comparative Education Review, 1974, 18, 164-179.

Pyrczak, F. and Axelrod, J. Determining the passage dependence of reading comprehension exercises. Journal of Reading, 1976, 19, 279-283.

Smith, R. J. and Barrett, T. C. Teaching reading in the middle grades. Reading, MA: Addison-Wesley Publishing Company, 1974.

Sokal, R. R. and Sneath, P. H. A. Principles of numerical taxonomy. San Francisco: W. H. Freeman, 1963.

Sörbom, D. and Jöreskog, K. G. Confirmatory factor analysis with model modification. Chicago: National Educational Resources, Inc., 1976.

Spache, G. D. and Spache, E. B. Reading in elementary school. Boston: Allyn and Bacon, 1973.

Thorndike, R. L. International studies in evaluation III: Reading comprehension education in fifteen countries: An empirical study. Stockholm: Almqvist and Wiksell, 1973.

Traxler, A. E. Values and limitations of standardized reading tests. In R. Farr (Ed.), Measurement and evaluation of reading. New York: Harcourt, Brace and World, 1970.

Tuinman, J. J. Determining the passage dependency of comprehension questions in five major tests. Reading Research Quarterly, 1973-74, 9, 206-223.

van Thillo, M. and Jöreskog, K. G. SIFASP (ETS RB 70-62). Princeton, N. J.: Educational Testing Service, 1970.

APPENDIX A

LAYOUT AND COMPUTATION OF THE TETRACHORIC CORRELATION COEFFICIENT

The SPSS program which computes tetrachoric correlation coefficients is called TETRACHORIC. TETRACHORIC first computes a 2 x 2 table such as the one shown below. A tetrachoric correlation is then computed from the proportion of cases falling into the various subcells of the table (p_a, p_b, p_c, and p_d).

		Variable 2: 1	Variable 2: 0
Variable 1	1	p_a	p_b
	0	p_c	p_d

From this table the following quantities are computed:

a) The marginal proportions: $p_1 = p_a + p_b$ and $p_2 = p_a + p_c$. The complementary proportions are $q_1 = 1.0 - p_1$ and $q_2 = 1.0 - p_2$.

b) The normal deviates corresponding to the proportions:

$$h = \Phi^{-1}(p_1), \text{ and } k = \Phi^{-1}(p_2).$$

Since the underlying distribution is assumed to be bivariate normal, the proportion d may be expressed in terms of h, k, and the unknown correlation ρ as:

$$d = \Phi(h,k;\rho) = 1/2\pi\sqrt{(1-\rho^2)} \int_{-\infty}^{k}\int_{-\infty}^{h} \exp\left[\frac{-(x^2 - 2xy + y^2)}{\sqrt{(1 - \rho^2)}}\right] dx\, dy$$

This equation is solved for ρ.

Reference Note: Matzek, M. A. Personal Communication, July, 1977.

APPENDIX B

TABLES OF UNROTATED FACTOR LOADINGS

Table B1

Unrotated Factor Loadings for Maximum Likelihood Factor Analysis of the IEA Reading Comprehension Test Items from England's Students

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
1	.369	.461	-.295	.178
2	.323	.509	-.085	.079
3	.396	.560	-.296	-.093
4	.318	.341	.128	-.227
5	.175	.256	-.257	.017
6	.197	.162	.170	-.255
7	.276	.385	.295	-.096
8	.361	.463	-.077	-.088
9	.468	.342	-.155	-.082
10	1.000	.000	.000	.000
11	.187	.292	.104	-.005
12	.622	.153	.119	-.047
13	.335	.483	.078	-.131
14	.367	.435	-.051	-.057
15	.250	.405	.136	-.055
16	.208	.474	.227	-.045
17	.017	-.017	.102	.157
18	.133	.136	.071	.136
19	.142	.274	.128	.227
20	.103	.182	.081	.108
21	.371	.421	-.154	.016
22	.264	.357	-.137	-.086
23	.448	.502	-.002	-.192
24	.442	.555	.017	-.205
25	.332	.467	.218	-.060
26	-.002	.033	-.045	.019

Table B1 (continued)

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
27	.290	.351	-.124	-.128
28	.287	.515	.039	-.030
29	.282	.335	-.167	-.053
30	.309	.459	-.110	-.062
31	.202	.441	.006	-.050
32	.233	.454	.013	-.002
33	.356	.439	-.024	-.024
34	.307	.315	.067	-.032
35	.308	.542	.097	.159
36	.256	.413	.165	.004
37	.127	.228	.076	-.048
38	.366	.485	.119	-.044
39	.405	.566	-.216	.091
40	.371	.534	.045	.136
41	.197	.255	-.013	.030
42	.220	.570	.112	.103
43	.351	.566	-.084	.068
44	.233	.457	.137	.130
45	.162	.368	.236	.087
46	.319	.479	.060	.243
47	.255	.372	-.034	.070
48	.313	.550	-.053	-.004
49	.153	.391	.010	.099
50	.234	.360	.121	.216
51	.066	.159	-.081	.013
52	.227	.300	.225	.030
Percent of Variance Explained by Each Factor	10.6	16.3	1.9	1.2

Table B2

Unrotated Factor Loadings for Maximum Likelihood Factor Analysis of the IEA Reading Comprehension Test Items from New Zealand's Students

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
1	.269	.589	.406	-.041
2	.201	.523	.112	-.121
3	.148	.509	.145	-.014
4	.203	.457	-.050	.222
5	.170	.303	-.035	.078
6	.222	.350	-.195	.116
7	.242	.491	-.060	.112
8	.354	.531	.109	-.052
9	.365	.510	.320	-.088
10	1.000	.000	.000	.000
11	.272	.397	-.067	.092
12	.572	.177	-.060	-.012
13	.192	.517	-.084	-.077
14	.227	.484	.141	.040
15	.133	.585	-.307	-.052
16	.203	.609	-.197	-.059
17	.053	.149	-.058	-.291
18	.105	.280	-.105	-.173
19	.113	.370	-.118	-.149
20	-.015	.300	-.166	-.145
21	.106	.571	.225	.144
22	.131	.414	.046	.114
23	.222	.656	.103	.181
24	.296	.649	-.139	.412
25	.168	.491	-.250	.060
26	.040	.081	.201	.002

Table B2 (continued)

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
27	.149	.469	.011	-.075
28	.164	.582	-.129	-.181
29	.129	.465	-.009	-.161
30	.188	.568	.107	-.049
31	.119	.410	-.034	-.072
32	.145	.479	.015	-.120
33	.175	.534	.228	.137
34	.109	.373	-.092	.079
35	.180	.483	-.115	-.009
36	.266	.430	-.039	.032
37	.123	.280	-.314	.058
38	.090	.440	-.116	-.034
39	.180	.632	.180	-.074
40	.188	.523	.060	-.165
41	.104	.205	.049	-.039
42	.185	.597	-.104	-.158
43	.140	.568	.119	.007
44	.161	.455	-.030	-.060
45	.083	.353	-.202	-.068
46	.056	.562	.002	-.159
47	.134	.366	.014	-.001
48	.132	.584	-.127	.033
49	.165	.308	.136	.037
50	.107	.375	-.064	-.249
51	.089	.120	.261	.059
52	.124	.432	-.283	-.118
Percent of Variance Explained by Each Factor	5.6	21.1	2.5	1.5

Table B3

Unrotated Factor Loadings for Maximum Likelihood Factor Analysis of the IEA Reading Comprehension Test Items from United States' Students

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
1	.318	.430	-.059	.275
2	.321	.488	-.032	.096
3	.380	.579	-.156	.066
4	.332	.460	-.204	-.214
5	.197	.348	-.006	-.086
6	.266	.241	-.226	-.165
7	.359	.476	-.116	-.137
8	.382	.471	-.084	.103
9	.321	.355	-.079	.015
10	1.000	.000	.000	.000
11	.226	.285	-.035	-.007
12	.596	.040	.034	.049
13	.266	.548	.000	-.075
14	.296	.486	-.138	.016
15	.265	.589	.090	-.154
16	.262	.588	.104	-.171
17	.023	.096	.196	.050
18	.075	.274	.025	.082
19	.239	.333	.240	.068
20	.079	.177	.256	.001
21	.315	.319	-.175	.235
22	.185	.352	-.128	.103
23	.397	.590	-.180	-.107
24	.409	.639	-.194	-.092
25	.228	.382	.032	-.165
26	.133	.108	-.029	.136

Table B3 (continued)

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
27	.281	.399	-.134	-.099
28	.271	.431	-.015	-.055
29	.229	.392	-.013	.093
30	.244	.393	-.040	.067
31	.187	.346	-.064	-.126
32	.286	.441	.102	-.040
33	.396	.521	-.221	.074
34	.301	.422	-.062	-.054
35	.295	.529	.115	-.072
36	.286	.434	.023	-.097
37	.169	.195	.119	-.175
38	.061	-.015	.066	-.186
39	.381	.558	.029	.195
40	.285	.504	.160	.057
41	.132	.203	.030	.038
42	.281	.559	.282	.026
43	.274	.534	.137	.168
44	.211	.372	.183	.036
45	.263	.521	.256	-.104
46	.267	.505	.080	.106
47	.234	.467	-.009	.199
48	.288	.572	.022	.073
49	.156	.281	.101	.051
50	.149	.391	.191	.119
51	.145	.252	-.008	.239
52	.294	.380	.278	-.205
Percent of Variance Explained by Each Factor	9.5	17.7	1.8	.5

APPENDIX C

TABLES OF ROTATED FACTOR LOADINGS 54

Table C1

Oblique Solutions for Maximum Likelihood Factor Analysis of the IEA Reading Comprehension Test Items from England's Students

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
1	.221	.063	.530	.094
2	.102	.331	.311	.044
3	.178	.244	.483	-.176
4	.138	.444	-.071	-.179
5	.097	-.009	.361	-.056
6	.091	.340	-.192	-.193
7	.048	.595	-.173	-.017
8	.159	.338	.228	-.106
9	.362	.124	.252	-.096
10	1.146	-.241	-.038	.127
11	.033	.334	.011	.017
12	.605	.118	-.093	.050
13	.091	.507	.069	-.112
14	.180	.320	.200	-.066
15	.034	.474	.007	-.025
16	-.066	.626	-.049	-.004
17	.024	.012	-.058	.183
18	.075	.116	.027	.154
19	.004	.269	.057	.245
20	.010	.187	.028	.121
21	.213	.195	.321	-.021
22	.117	.204	.247	-.122
23	.218	.452	.132	-.180
24	.178	.527	.132	-.194
25	.080	.587	-.053	-.004
26	-.013	-.011	.065	.003

Table C1 (continued)

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
27	.146	.218	.216	-.156
28	.032	.485	.156	-.034
29	.157	.143	.278	-.094
30	.109	.310	.269	-.096
31	-.021	.411	.155	-.066
32	.009	.407	.168	-.015
33	.163	.338	.186	-.029
34	.162	.306	.044	-.005
35	.044	.493	.172	-.165
36	.036	.486	.001	.039
37	.003	.275	-.001	-.033
38	.125	.508	.057	-.013
39	.185	.253	.466	.025
40	.127	.434	.210	.137
41	.086	.185	.122	.025
42	-.078	.573	.153	.101
43	.101	.385	.330	.032
44	-.003	.469	.091	.147
45	-.052	.495	-.057	.131
46	.103	.369	.210	.244
47	.093	.257	.203	.054
48	.057	.428	.270	-.034
49	-.039	.327	.181	.078
50	.062	.331	.094	.236
51	.000	.069	.149	-.017
52	.059	.421	-.096	.088

Table C2

Oblique Solutions for Maximum Likelihood Factor Analysis of the IEA Reading Comprehension Test Items from New Zealand's Students

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
1	.101	.006	.175	.639
2	.079	.059	.319	.306
3	.016	.147	.199	.347
4	.099	.452	-.041	.146
5	.109	.228	.041	.085
6	.169	.367	.055	-.059
7	.139	.361	.090	.136
8	.243	.113	.220	.304
9	.240	-.047	.190	.509
10	1.106	-.128	-.196	-.056
11	.199	.293	.061	.087
12	.594	.037	.003	-.023
13	.089	.206	.322	.106
14	.107	.180	.114	.333
15	.029	.391	.398	-.090
16	.089	.328	.375	.028
17	.041	-.189	.391	-.031
18	.062	.012	.332	-.017
19	.048	.088	.351	.008
20	-.070	.099	.350	-.062
21	-.064	.298	.047	.473
22	.026	.281	.043	.220
23	.053	.428	.058	.379
24	.151	.773	-.146	.146
25	.081	.420	.210	-.061
26	.001	-.069	-.023	.238

Table C2 (continued)

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
27	.044	.138	.280	.187
28	.052	.162	.484	.077
29	.030	.064	.380	.157
30	.050	.157	.268	.326
31	.032	.139	.265	.118
32	.039	.096	.333	.191
33	.022	.262	.021	.457
34	.029	.302	.101	.059
35	.058	.273	.243	.067
36	.184	.236	.136	.124
37	.094	.351	.134	-.210
38	-.001	.239	.266	.052
39	.017	.127	.308	.426
40	.072	.044	.383	.248
41	.058	.026	.109	.124
42	.067	.177	.455	.111
43	-.008	.213	.213	.347
44	.064	.167	.265	.140
45	.025	.208	.284	-.076
46	-.078	.118	.437	.213
47	.050	.159	.152	.156
48	.005	.379	.259	.103
49	.085	.100	.044	.260
50	.039	-.037	.449	.056
51	.036	-.032	-.090	.319
52	.060	.237	.390	-.134

Table C3

Oblique Solutions for Maximum Likelihood Factor Analysis of the IEA Reading Comprehension Test Items from United States' Students

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
1	.158	.284	.092	.323
2	.095	.428	.096	.155
3	.081	.614	-.021	.173
4	.024	.695	-.162	-.098
5	.001	.387	.055	-.037
6	.079	.466	-.228	-.080
7	.076	.604	-.048	-.051
8	.161	.437	.035	.166
9	.140	.374	-.006	.065
10	1.068	-.079	-.058	-.132
11	.080	.295	.023	.030
12	.632	-.059	.023	-.034
13	-.030	.571	.113	.002
14	.033	.544	-.032	.112
15	-.056	.599	.202	-.090
16	-.061	.599	.213	-.110
17	.016	-.066	.244	.008
18	-.047	.204	.112	.116
19	.129	.115	.345	.031
20	.030	.000	.316	-.047
21	.188	.273	-.069	.297
22	.009	.362	-.031	.181
23	.051	.741	-.082	.014
24	.038	.790	-.082	.040
25	.004	.440	.085	-.121
26	.109	.038	.019	.142

Table C3 (continued)

Item Number	Factor 1	Factor 2	Factor 3	Factor 4
27	.040	.524	-.075	-.015
28	.041	.452	.070	.001
29	.051	.328	.096	.140
30	.056	.362	.062	.122
31	-.027	.448	-.015	-.058
32	.073	.375	.199	-.017
33	.121	.594	-.105	.185
34	.070	.471	.016	.011
35	.030	.471	.229	-.035
36	.052	.452	.101	-.050
37	.055	.205	.130	-.185
38	.042	.046	.022	-.208
39	.153	.395	.199	.239
40	.070	.343	.299	.072
41	.045	.150	.086	.051
42	.049	.335	.437	.018
43	.062	.323	.308	.192
44	.061	.213	.289	.028
45	.017	.391	.372	-.103
46	.049	.368	.226	.144
47	.040	.338	.143	.255
48	.017	.492	.174	.139
49	.042	.179	.184	.055
50	.005	.182	.324	.117
51	.071	.105	.102	.262
52	.107	.292	.335	-.239

APPENDIX D

TREE DIAGRAMS

Complete Linkage Cluster Analysis of the IEA Reading Comprehension Test Items for Students from England

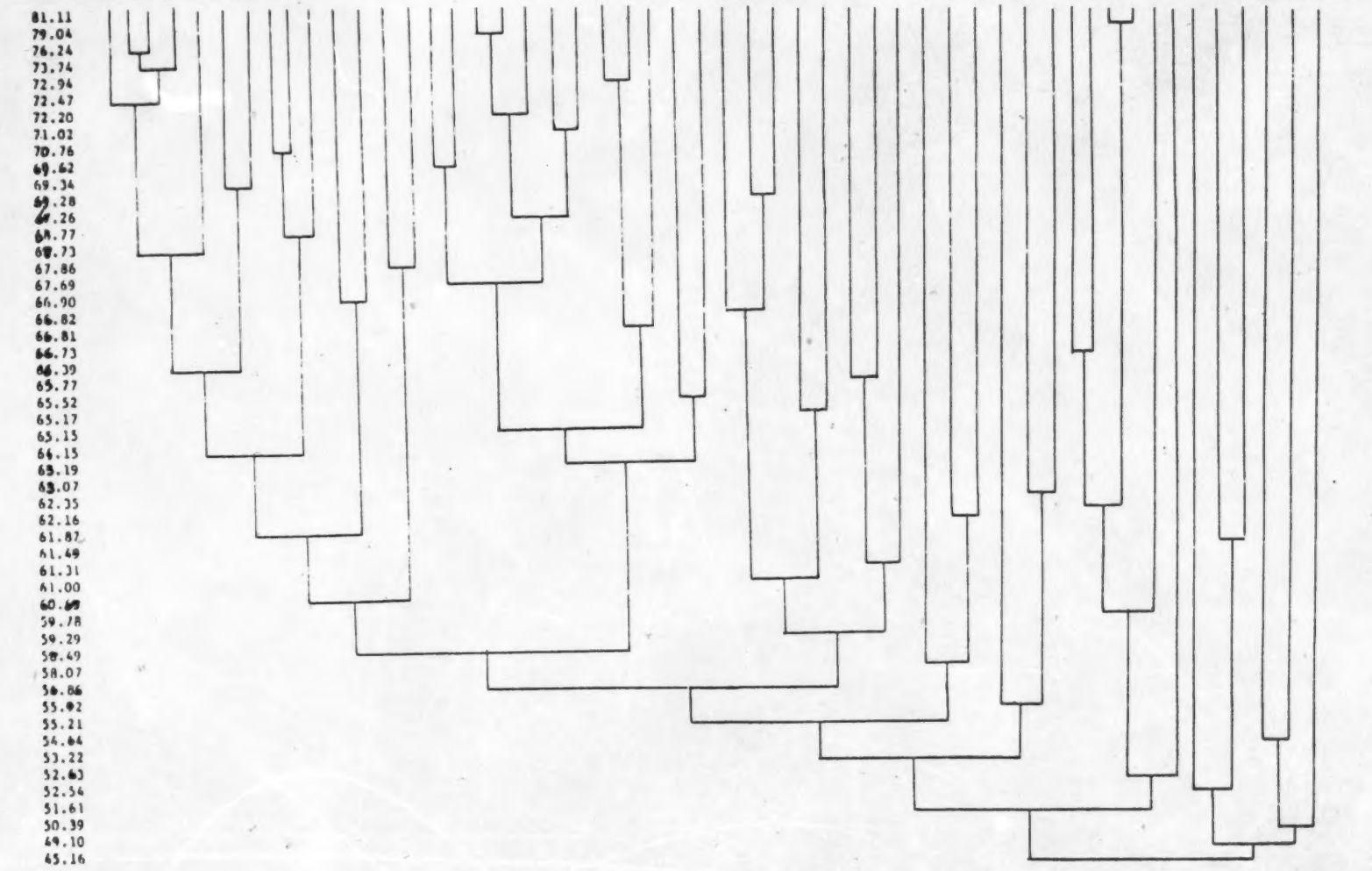

Figure D1

Complete Linkage Cluster Analysis of the IEA Reading Comprehension Test Items for Students From New Zealand

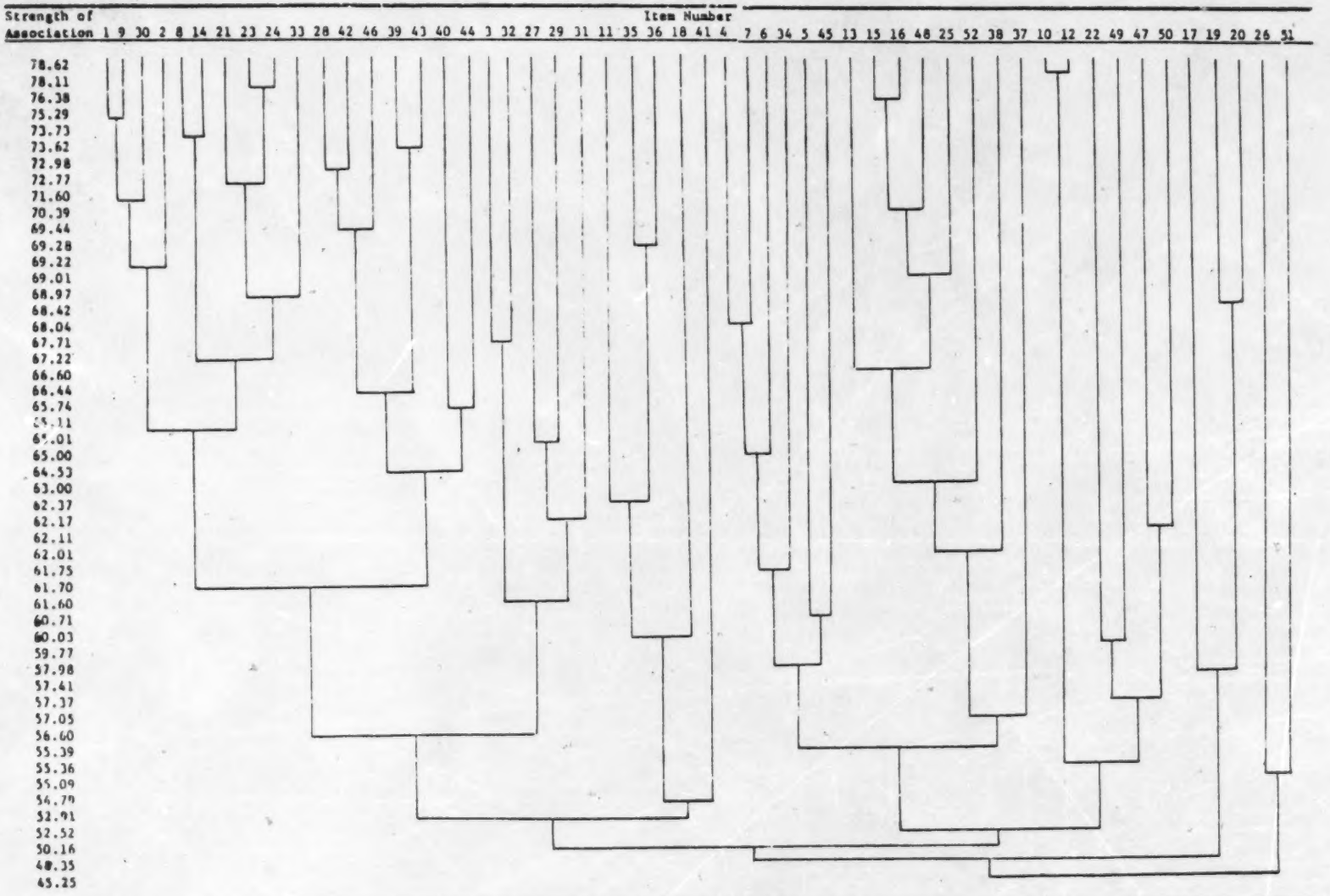

Figure D2

Complete Linkage Cluster Analysis of the IEA Reading Comprehension Test Items for Students from United States

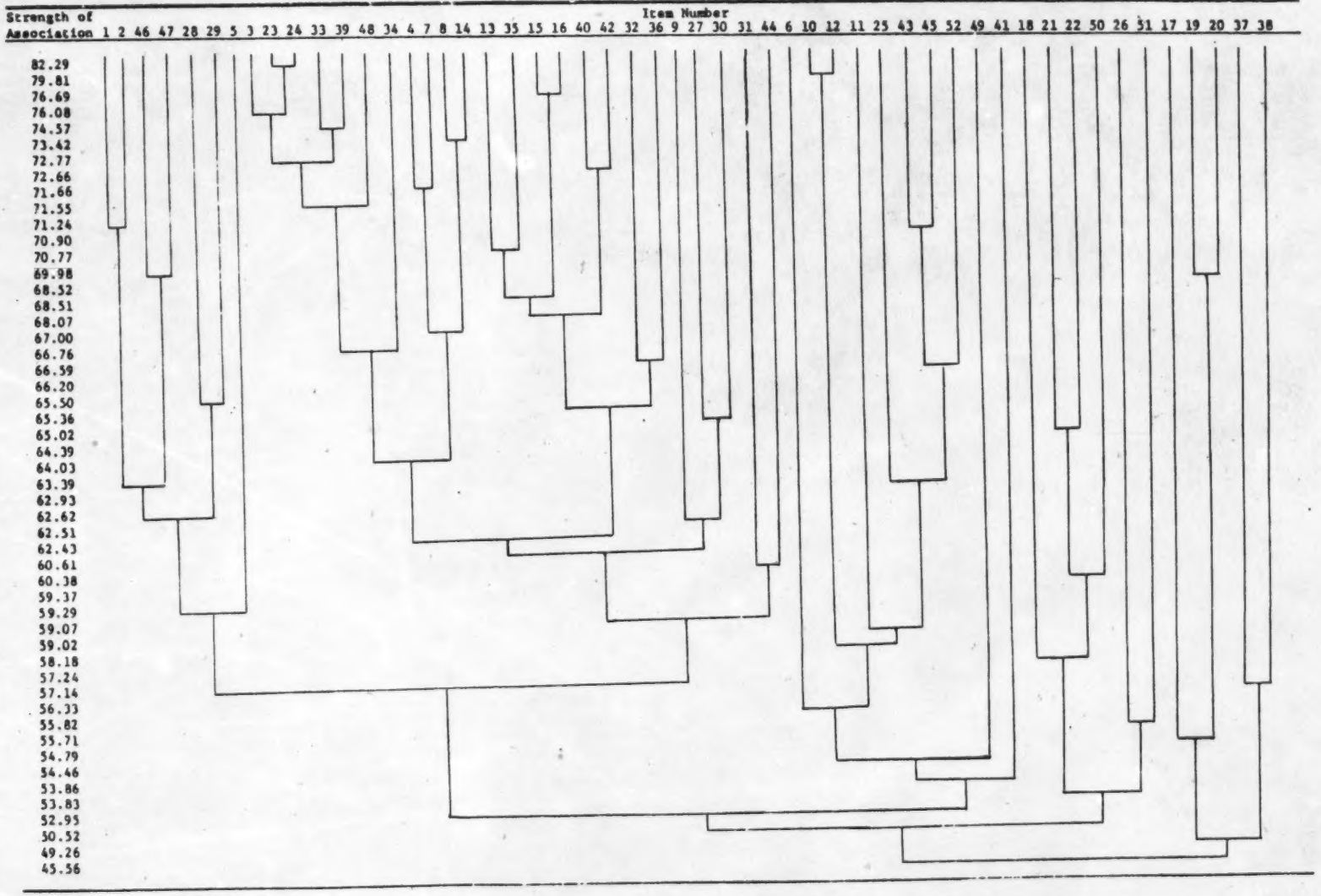

Figure D3